THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

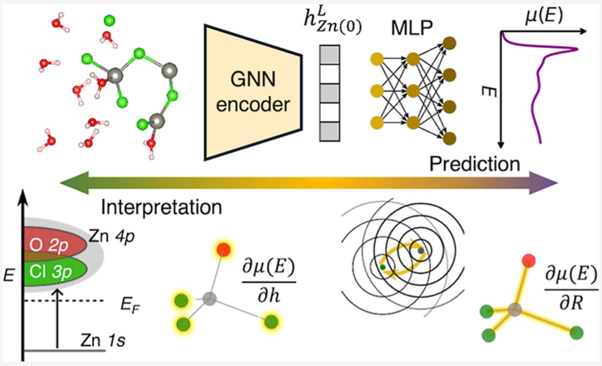
Deciphering the Solvation Structure of Aqueous $ZnCl_2$ Solutions from X-ray Absorption Spectra Using the Interpretable Graph Neural Network

Chuntian Cao,* Boyang Li, Armando Rodriguez Campos, Alexis Pace, Joshua J. Kas, Xifan Wu, Lu Ma, Dali Yang, Wei Xu, Shinjae Yoo, Esther S. Takeuchi, Kenneth J. Takeuchi, Shan Yan, Amy C. Marschilok,* and Deyu Lu*

Cite This: https://doi.org/10.1021/acs.jpcb.6c00464

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Machine learning (ML) provides powerful pathways for predicting spectroscopic observables from atomic structures, but its broader impact depends on making model predictions interpretable in terms of physical and chemical principles. Here, we introduce a physics-guided graph neural network (GNN) model that predicts Zn K-edge X-ray spectroscopy (XAS) spectra of aqueous $ZnCl_2$ solutions. Training data are generated from ab initio XAS calculations on molecular dynamics snapshots obtained using a machine learning interatomic potential. The GNN reproduces experimental spectra across concentrations from dilute (<0.1 m) to highly concentrated (30 m, "water-in-salt") regimes and scales efficiently to large, disordered liquid systems beyond the reach of conventional ab initio approaches. Gradient-based attribution analysis reveals that the model learns physically meaningful structure-spectrum relationships. Ligand-specific attributions reflect orbital hybridization patterns and the origin of the excitations derived from the density functional theory. Bond-length attributions recover spectral shifts consistent with multiple-scattering theory. This work bridges data-driven prediction with electronic-structure theory, establishing a general paradigm for interpretable ML that links atomic structure, electronic structure, and spectroscopic observables.

1. INTRODUCTION

X-ray absorption spectroscopy (XAS) provides detailed, element-specific information on local atomic and electronic structures, including coordination number, ligand type, bond length, oxidation state, and orbital hybridization[1−3] and is widely used to gain mechanistic understanding of complex chemical environments.[4,5] Accurate and efficient simulation of spectroscopic data is critical for interpreting experimental spectra and conducting rational materials design. Conventional ab initio approaches can accurately simulate XAS, but they are computationally demanding for large systems. Machine learning (ML), as an emerging data-driven technique, offers a promising avenue to learn structure-spectrum relationships directly from data and has been applied to predict spectral functions in XAS,[6−18] as well as Raman and infrared spectroscopy.[19−22]

Despite the rapid progress of ML, its widespread adoption in materials research has been hindered by the perceived "black box" nature of typical off-the-shelf models and the lack of interpretability in terms of physical and chemical principles. Recent advances in explainable ML have begun to bridge data-driven predictions and physical understanding in materials chemistry.[23−25] In the context of XAS, Kotobi et al. combined class activation maps (CAM) with time-dependent density functional theory (DFT) ground truth orbital analysis in small molecular systems in order to quantitatively validate model attributions based on physical orbital contributions.[9] This study highlights the promise of interpretable ML for spectroscopy while also motivating the extensions to more complex and disordered systems.

In this study, we develop a graph neural network (GNN)-based ML model to predict Zn K-edge XAS for aqueous $ZnCl_2$ solutions from dilute (<0.1 m) to highly concentrated (30 m, "water-in-salt") regimes. $ZnCl_2$ electrolytes are promising aqueous electrolytes for next-generation Zn-ion batteries.[26,27] Understanding the solvation structure and ion−ligand interactions is important for optimizing the electrochemical

Received: January 22, 2026
Revised: April 7, 2026
Accepted: April 20, 2026

ACS Publications

stability and ion transport properties of battery electrolytes. In a recent study, Driscoll et al. used XAS to investigate the Zn^{2+} solvation structure in $ZnCl_2$ aqueous solutions.[28] The qualitative spectral trend was reproduced through ab initio simulations based on cluster models of selective solvation motifs. This study revealed that tetrahedral $ZnCl_4^{2-}$ species exhibit a lower energy white line than do octahedral $Zn(H_2O)_6^{2+}$ complexes. Beyond the correct qualitative trend, a more detailed study is needed to correlate the spectral features with a wealth of solvation motifs and identify their electronic structure origins. However, obtaining the in-depth understanding is challenging, because of the roadblock to simulate the dynamics and compute the XAS spectra of liquid systems containing thousands of atoms in the supercell.

Here, we address this challenge using large-scale molecular dynamics (MD) simulations and interpretable ML. The training data consists of atomic configurations generated by MD simulations using a machine learning interatomic potential (MLIP) developed in our previous work that can accurately reproduce the DFT potential energy surface of $ZnCl_2$ solution at a wide range of concentrations.[29] XAS spectra are generated by using the ab initio core-hole potential method on MD snapshots. The XAS ML model uses a GNN encoder to extract absorber feature embeddings from atomic structures. Subsequently, a multilayer perceptron (MLP) spectral head predicts site-specific spectrum from these latent embeddings. The design is physics-guided because unlike total energy prediction models which stack latent embeddings from all atoms, our model focuses on the absorbing atom, passing only its latent embedding to the spectrum head. This aligns with XAS's nature as a site-specific local probe and enhances model interpretability. The ML model accurately reproduces experimental spectra across all concentrations studied.

To interpret the learned structure−spectrum relationships, we first use Uniform Manifold Approximation and Projection (UMAP)[30] to visualize the latent feature space and confirm that the GNN encodes the local coordination environment of the absorber. We then use the Integrated Gradients (IG) method[31,32] to quantify the influence of atom representation and bond length on the predicted spectra. IG analyses reveal that ligand-specific attributions correlate with orbital hybridization patterns in the unoccupied electronic states. Bond-length-dependent attributions capture spectral shifts, consistent with multiple-scattering theory. By connecting ML attributions to established spectroscopic physics, this work demonstrates that data-driven models can capture physically meaningful principles. This study establishes a generalizable pathway for interpretable ML in spectroscopy that bridges atomic structure, electronic structure, and a measurable spectral response.

2. METHODS

2.1. Experimental XAS Measurements

$ZnCl_2$ solutions were prepared by dissolving anhydrous $ZnCl_2$ (Sigma-Aldrich, ≥98%) in 18.2 MΩ DI water and then stirring overnight with a sealed vial. Solutions were transferred into plastic pouches individually and sealed for XAS data collection. The XAS measurements were performed at the 7-BM (QAS) beamline in the National Synchrotron Light Source (NSLS-II) at Brookhaven National Laboratory. XAS measurements were acquired in transmission mode at the Zn−K edge along with the reference foil for each sample. Four scans were collected for each sample and then merged to improve the signal-to-noise ratio. The energy edge was then calibrated by Zn reference foil at 9659 eV. All the calibration, merging, and normalization process were done using Athena software from the Demeter package.[33] Our measured Zn K-edge XAS spectra are in good agreement with literature[34] except for a constant energy shift due to different treatments of edge alignment.

2.2. Generation of Atomic Configurations

Atomic configurations of aqueous $ZnCl_2$ solutions were generated by MD simulations using an MLIP developed in our previous work.[29] The MLIP was trained with ab initio data obtained from Quantum Espresso[35,36] using the SCAN[37] exchange−correlation functional and implemented through the DeePMD-kit[38] framework. The cutoff radius was set to 6 Å. The MLIP predicts the energy of the system and forces on each atom from atomic coordination and species. MD simulations were performed in LAMMPS[39] under isothermal−isobaric (NpT) ensemble at 333 K and 1 bar. At 333 K, the SCAN functional is known to reproduce the structural properties of liquid water near room temperature.[40,41] The Nosé−Hoover thermostat and barostat as implemented in LAMMPS were employed. The thermostat relaxation time was set to 0.05 ps, and the barostat relaxation time was set to 0.5 ps.

For training data generation, simulation cells of approximately 10 Å × 10 Å × 10 Å were used with periodic boundary conditions. Snapshots were extracted from trajectories at four concentrations (1.85, 3.5, 18.5, and 30 m). An additional test data set corresponding to a Zn/O ratio of 1:7 (8 m) was generated to evaluate the model's interpolation capability.

For production simulations used to compare with experimental XAS data, larger systems were constructed with approximately 25 Å × 25 Å × 25 Å (about 1600 atoms). For the dilute 0.1 m system, the production run system consists of about 6000 atoms to ensure an accurate representation of solvation environments. In each MD simulation, the first 2 ns were considered as equilibration time and were discarded. Production trajectories were recorded, with snapshots saved every 5 ps. Structural configurations for spectra prediction were randomly sampled from the MD trajectory. Detailed composition information about the simulation systems is provided in Table S1.

2.3. Ab Initio XAS Calculations

XAS were calculated using the core-hole potential method[42] implemented in VASP.[43,44] The input files were generated using the Lightshow package.[45] The energy levels and wave functions were obtained using DFT, and the Perdew−Burke−Ernzerhof (PBE)[46] generalized gradient approximation was used for the exchange−correlation functional. The GW-type projector augmented-wave (PAW) pseudopotentials were used to ensure accurate treatment of unoccupied states relevant to XAS transitions. The simulation was carried out using the excited-electron and core-hole (XCH) approximation.[47] The final training data set comprised approximately 2000 Zn site spectra paired with their corresponding atomic configurations extracted from MD simulations.

Each simulated spectrum was broadened by convolution with a Lorentzian function of 1.67 eV half-width at half-maximum, corresponding to the Zn 1s core-hole lifetime.[48] Simulated spectral alignments[49] across configurations were performed by shifting the calculated edge according to $E_{\mathrm{align}} = (E - E_{\mathrm{Fermi}}) + (E_{\mathrm{XCH}} - E_{\mathrm{GS}})$, where E_{XCH} is the total energy of the excited state (core-hole system), E_{Fermi} is the Fermi energy of the core-hole system, and E_{GS} is the total energy of the ground state. This alignment shifts the energy scale so that the absorption energy corresponds to the difference between the total energies of the excited and ground states. A constant global energy shift was then applied to match the experimental reference spectrum. For intensity normalization, the predicted spectra were scaled to match the normalized experimental spectra. Specifically, predicted spectra for concentrations from 0.1 to 30 m were compared with the corresponding experimental spectra, and the postedge region from 9675 to 9715 eV was used to determine a scaling factor. Then the average scaling factor over all concentrations was applied uniformly to all of the predicted spectra.

The PDOS was computed by using the same pseudopotential in the XCH approach as described for the XAS simulations. All of the

https://doi.org/10.1021/acs.jpcb.6c00464
J. Phys. Chem. B XXXX, XXX, XXX−XXX

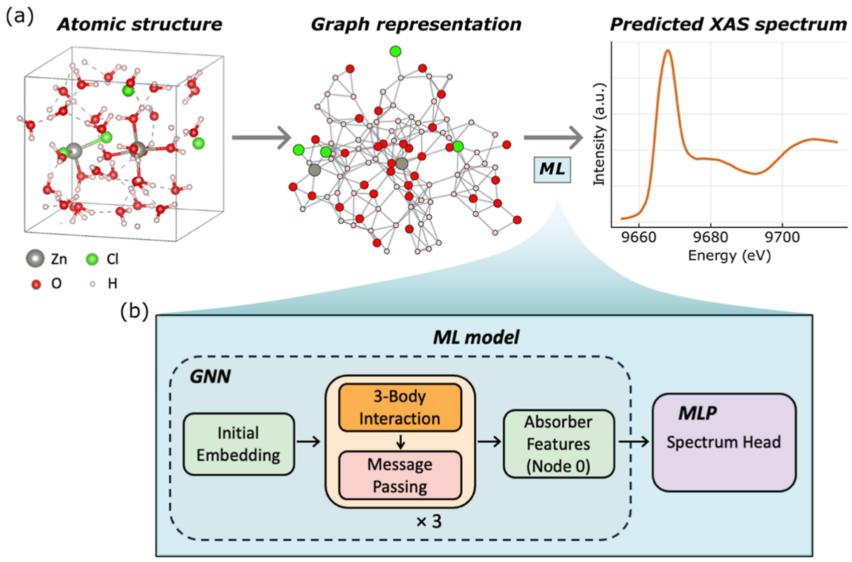


Figure 1. Illustration of the workflow and machine learning (ML) model. (a) Atomic structures are converted into graph representations, and the trained ML model predicts XAS from the graph. Here, the graph representation is for illustration only, and not all edges are presented. (b) Architecture of the ML model. It consists of a graph neural network (GNN) encoder that extracts absorber features, followed by a multilayer perception (MLP) spectrum head that maps these features to the predicted XAS. The GNN encoder includes initial node and edge embeddings, three message-passing layers, and three-body interaction layers to capture angular correlations.

PDOS results are plotted relative to the Fermi level. The alignment of the PDOS is the same as that of the XAS spectra, except for a constant corresponding to the 1s core level from the XCH pseudopotential.

2.4. Many-Body Shakeup Correction

Many-body shakeup effects were incorporated by convolving the calculated XAS spectrum with a core-hole spectral function $A(\omega)$.[50−53] The spectral function $A(\omega) = -\frac{1}{\pi}\,\mathrm{Im}[G(\omega)]$ was calculated within the cumulant approximation of the core-hole Green's function, $G(t) = G_0(t)\,\exp(C(t))$, where G_0 is the single-particle Green's function and $C(t)$ is the cumulant function, which incorporates the many-body excitations[54,55] induced by the sudden creation of the core hole.

In practice, the spectral function $A(\omega)$ was computed for representative Zn coordination environments, including $Zn(H_2O)_6$, $Zn(H_2O)_5Cl$, $Zn(H_2O)_4Cl$, $Zn(H_2O)_3Cl_2$, $Zn(H_2O)_2Cl_2$, $Zn(H_2O)Cl_3$, and $ZnCl_4$. The resulting spectral functions were found to have similar satellite intensities, and convolution results using different spectral functions show minimal differences (Figure S1). Therefore, an average $A(\omega)$ was used for all spectra in the shakeup correction.

The convolution redistributes spectral weight from the quasiparticle peak to shakeup satellites associated with charge-transfer and plasmon excitations. This correction effectively reduces the white-line intensity and fine-structure amplitude, improving the agreement between the calculated and experimental XAS spectra by accounting for many-body inelastic losses not captured in single-particle DFT calculations.

2.5. ML Model

Each atomic configuration is represented as a graph $G = (V, E)$, where $V = \{v_i\}$ denotes the set of atomic nodes, and $E = \{(v_i, v_j)\}$ denotes edges connecting neighboring atoms within a cutoff radius $r_{\mathrm{cut}} = 4$ Å. The graph is used as the input to the GNN. The GNN follows the design and architecture of M3GNet,[56] simplified and specialized for our system of four element types (Zn, O, H, and Cl) and for predicting spectral curves rather than scalar total energies. The GNN consists of node and edge embeddings, three-body interactions, and three message passing layers with a cutoff distance of 4 Å for neighbor communication. After the message-passing layers, the latent feature vector of the absorbing atom (Zn) is extracted and serves as the input to the spectrum head. The spectrum head is implemented as a feed-forward MLP with two hidden layers (64 → 128 → 128 → 200 neurons) and SiLU activation. The MLP outputs the predicted absorption spectrum $\mu(E)$ sampled at 200 energy points. The network is trained using the MSE loss between predicted and reference spectra.

2.6. Integrated Gradients (IG) Attribution

The IG method[31,57] implemented using the Captum library[58] is employed to analyze how structural features influence the predicted XAS spectra. IG quantifies how each input feature contributes to the model output by integrating the gradients along a continuous path from a chosen baseline input to the actual input.

For node-wise attribution, the Integrated Gradient for atom i is the path integral of the gradient of the predicted intensity $\mu(E)$ with respect to the atom's initial embedding $h_i^{(0)}$:

$$\mathrm{IG}_{i,\mathrm{d}}(E) = (h_{i,\mathrm{d}}^{(0)} - h_{\mathrm{d}}^{\mathrm{base}}) \int_0^1 \frac{\partial \mu(E, h^{\mathrm{base}} + \alpha(h_i^{(0)} - h^{\mathrm{base}}))}{\partial h_{i,\mathrm{d}}^{(0)}}\, \mathrm{d}\alpha$$

where $h_i^{(0)}$ is the 64-dimensional initial embedding of node i, h^{base} is the baseline embedding, and the integral is evaluated along the linear path between them. $\mathrm{IG}_{i,\mathrm{d}}(E)$ represents the integrated contribution for the dth embedded feature of node i to the predicted intensity at energy E.

For the standard importance attribution, the baseline is a zero vector, corresponding to the absence of structural information. Since the sign of the embedding has no physical meaning, the node attribution is computed as the L_1-norm of the integrated gradients across all embedding dimensions:

$$|\mathrm{Attr}_i(E)| = \sum_{d=1}^{64} |\mathrm{IG}_{i,d}(E)| \quad \text{(with zero baseline)}$$

For the counterfactual attribution, the baseline is defined as the chemically neutral reference, which lies midway between the O and Cl embedding vectors: $h^{\mathrm{base}} = \frac{1}{2}(h_{\mathrm{O}}^{(0)} + h_{\mathrm{Cl}}^{(0)})$. In this case, the sign of each $\mathrm{IG}_{i,\mathrm{d}}(E)$ is physically interpretable. The CF attribution is therefore computed as

$$\mathrm{Attr}_i^{\mathrm{CF}}(E) = \sum_{d=1}^{64} \mathrm{IG}_{i,d}(E) \quad \text{(with O−Cl baseline)}$$

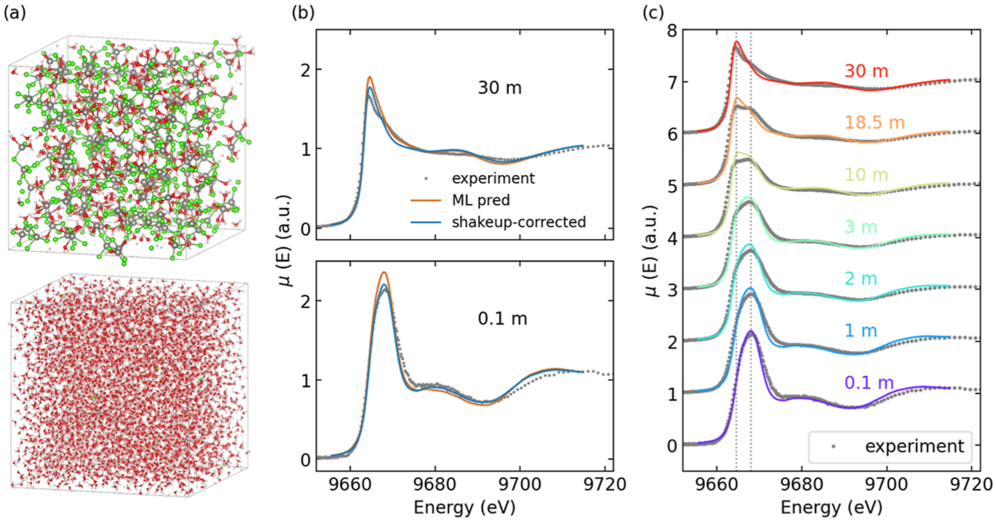


Figure 2. Comparison of ML-predicted and experimental XAS spectra. (a) MD snapshots of $ZnCl_2$ aqueous solutions. Top: concentrated 30 m system containing 200 $ZnCl_2$ and 368 H_2O molecules. Bottom: dilute 0.1 m system containing 2220 H_2O and 4 $ZnCl_2$ molecules. (b) Effect of the post hoc many-body shakeup correction on the predicted spectra. Top: 30 m; bottom: 0.1 m. (c) ML-predicted Zn K-edge XAS spectra at different concentrations (colored lines) after applying the shakeup correction, compared with experimental measurements (gray markers).

For bond-length attribution, the IG value for the ith bond is calculated as the path-integrated gradient of the predicted intensity $\mu(E)$ with respect to bond length R_i:

$$\mathrm{IG}_i(E) = (R_i - R_i^{\mathrm{base}}) \int_0^1 \frac{\partial \mu(E, R_i^{\mathrm{base}} + \alpha(R_i - R_i^{\mathrm{base}}))}{\partial R_i}\, \mathrm{d}\alpha$$

where R_i is the original length of bond i, and R_i^{base} is the baseline used to define the integration path, chosen as a small elongation of the corresponding bond, $R_i^{\mathrm{base}} = R_i + 0.05$ Å. Since $R_i - R_i^{\mathrm{base}} < 0$, we take the negative of integrated gradients as the bond length attribution:

$$\mathrm{Attr}_i^{\mathrm{bond\_len}}(E) = -\mathrm{IG}_i(E)$$

which corresponds to the change in spectral intensity upon increasing the bond length. This enables direct interpretation of derivative $\partial\mu(E)/\partial R_i$ as the spectral sensitivity to bond stretching.

3. RESULTS AND DISCUSSION

3.1. Model Performance and Experimental Validation

Figure 1 shows an illustration of the workflow and the GNN-based ML model. The model consists of two parts. The first part is a GNN encoder that transforms the atomic graph into a latent embedding of the absorbing Zn atom, termed the absorber features. The second part is an MLP spectrum head, which takes the absorber features as input and predicts the XAS spectrum. The GNN includes node and edge embeddings, three-body interactions, and three message passing layers. The ML model follows the architecture of M3GNet,[56] but is simplified and specialized for our system of four element types (Zn, O, H, and Cl) and for predicting spectral curves rather than the total energy; additionally, only the absorber features rather than the latent embeddings of all atoms are passed to the spectrum head for spectral prediction.

The ML model achieves good agreement between the predicted and simulated spectra. The mean squared error (MSE) of the predicted spectra is 2.57×10^{-3} on the validation set and 2.73×10^{-3} on the test set. The validation set is used to monitor model performance during training and prevent overfitting, and the test set contains unseen data used only for the final evaluation of the trained model. Figure S2 shows the decile plots of predictions on the test data set. Importantly, the test data set has a different concentration (8 m) than the training and validation data (1.85, 3.5, 18.5, and 30 m; see Table S1). Though similar atomic configurations may be included when sampling different concentrations, the good performance on the test set still demonstrates the model's strong interpolation capability.

To compare with experimental data, the trained model was applied to large-scale production systems containing about 1600 atoms, which represent more realistic experimental conditions. For a dilute concentration at 0.1 m (Zn/O = 1:555), a system with 2220 H_2O and 4 $ZnCl_2$ was used. This large system size underscores the critical role of the ML XAS surrogate, while the problem is otherwise intractable by using standard ab initio methods. Figure 2a shows two exemplary MD snapshots for concentrated and dilute systems. For 0.1−2 m concentrations, 10 snapshots were used to predict the XAS spectra. For 3−30 m concentrations, five snapshots were sufficient due to the larger number of Zn^{2+} sites in the simulation box. Details about the production systems, including the unit cell composition and the number of Zn absorber sites contributing to the averaged predicted spectrum, are provided in Table S1. For the very dilute 0.1 m system, the averaged spectrum is obtained from 40 Zn sites. The coordination statistics of O and Cl atoms in the FSS of Zn are shown in Figure S3 for both the selected snapshots and the full MD trajectory, demonstrating that the sampled snapshots provide a representative sampling of the FSS environment.

A many-body shakeup correction was applied by convolving the predicted spectra with the Zn 1s core-hole spectral function.[50−53] The shakeup correction was found to lower the white-line intensity and improve agreement with the experiment, as shown in Figure 2b. The ML-predicted spectra reproduce the experimental trends well (Figure 2c). At low concentration (0.1 m), a single white-line peak appears at 9668 eV. With increasing concentration, the peak intensity decreases and shifts slightly to lower energy, while a second peak emerges at lower energy at 9664.5 eV and becomes dominant at high concentration (30 m). This trend is also shown in Figure S4. Table S2 shows the root-mean-square error

https://doi.org/10.1021/acs.jpcb.6c00464

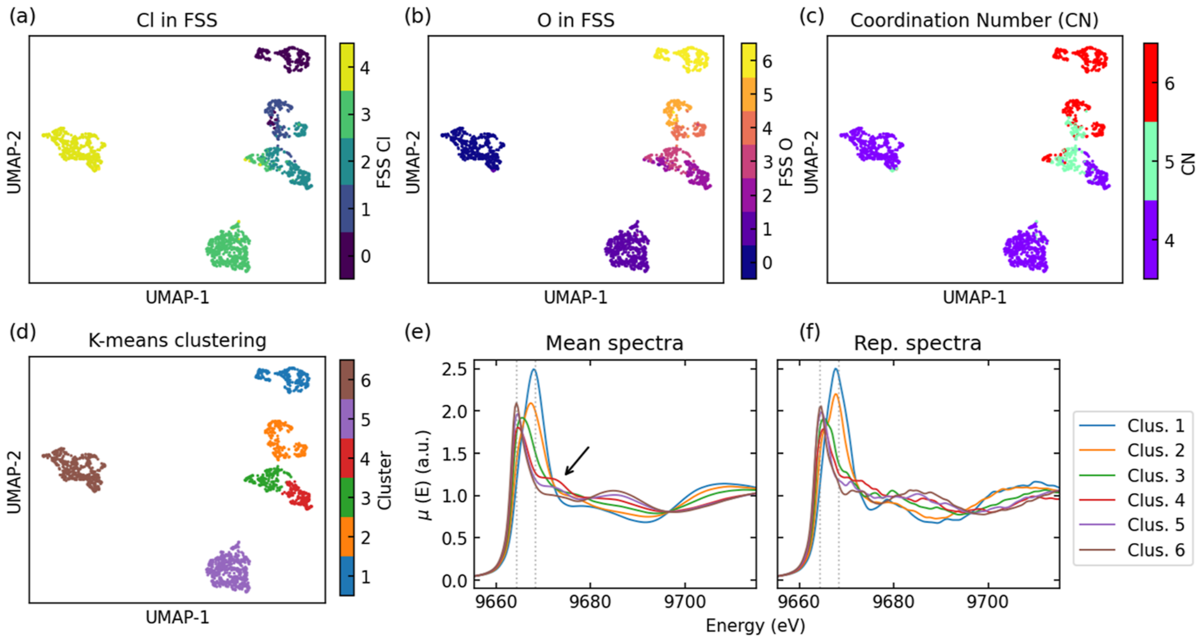


Figure 3. UMAP projection of the latent space features extracted by the GNN encoder. These absorber features serve as inputs to the spectrum head MLP. Panels (a−d) show the same UMAP distribution with different color codes. (a) Number of Cl ions in the first solvation shell (FSS) of the absorption Zn site. (b) Number of O atoms from water molecules in the FSS. (c) Total coordination number (CN) within the FSS of the absorbing Zn atom. The cutoff radius to define the FSS is in (a−c) is 3.0 Å. (d) Cluster assignments from *k*-means clustering. (e) Mean DFT-simulated spectra averaged within each of the six clusters. (f) Representative simulated spectra closest to the mean spectra.

(RMSE) and integrated absolute error (IAE) between the predicted and measured spectra.

To summarize, the ML model shows good interpolation capability as it accurately predicts spectra in the test set at concentrations not included in training. At dilute concentrations such as 0.1 m, ab initio simulations are computationally prohibitive due to the need for very large supercells to accurately sample the distribution of solvation motifs. In contrast, the ML model scales favorably for large systems, enabling reliable prediction of XAS spectra for dilute solutions in good agreement with experiment.

3.2. UMAP and Clustering Analysis of Learned Representations

The GNN encoder transforms the atomic graph into a latent representation of absorber features from which the subsequent spectrum head accurately predicts the spectrum. In this section, we analyze the latent absorber feature space to gain physical insight into what structural information the GNN has learned.

UMAP is applied to project the 64-dimensional absorber features into two dimensions for visualization and interpretation. As shown in Figure 3a−c, the 2D UMAP embedding exhibits clear clustering behavior. When color-coded by the number of Cl neighbors, number of O neighbors, and total coordination number (CN) within the first solvation shell (FSS) of the absorbing Zn atom, the UMAP distribution shows good correlation with the local coordination environment. Figure S5 shows that subtle variations in second-shell coordination and Zn−Zn neighbor counts also contribute to the latent-space data distribution. These results demonstrate that the latent features effectively embed the absorber environment with physical interpretability.

To analyze trends in the feature clusters, we applied the k-means[59] clustering algorithm to the 2D UMAP embedding. K-means is an unsupervised clustering method suitable for spherical clusters employed here to assign labels to the separated regions observed in the UMAP projection. Six clusters were selected based on physical intuition on the separation of first-shell species (O vs Cl) and coordination numbers. The resulting clusters are shown in Figure 3d, and the structural statistics are summarized in Table 1. UMAP with different initializations results in qualitatively consistent clustering assignments, as shown in Figure S6. The clusters closely align with the first-shell coordination environment. For example, cluster 1 corresponds to Zn sites coordinated by six O atoms and no Cl atoms, whereas cluster 2 contains Zn sites with four or five O and one or two Cl neighbors. Figure 3e shows the averaged simulated spectra from each cluster, and Figure 3f shows the representative simulated spectra, which have the smallest MSE from the averaged spectra with their

Table 1. Structural Statistics of the Six Clusters Shown in Figure 3d[a]

	FSS Cl		FSS O		CN		Rep. spectra	
cluster	mean	std	mean	std	mean	std	FSS Cl	FSS O
1	0	0	6	0	6	0	0	6
2	1.20	0.59	4.55	0.53	5.75	0.44	1	5
3	2.47	0.69	2.70	0.49	5.17	0.53	3	2
4	2.01	0.18	2.03	0.19	4.04	0.20	2	2
5	3.01	0.10	1.01	0.09	4.02	0.13	3	1
6	4.01	0.13	0.01	0.11	4.02	0.17	4	0

[a] The average local coordination environments are listed, including the number of Cl and O neighbors in the first solvation shell and the total coordination number. The last two columns list the local coordinates of the representative structures corresponding to the spectra shown in Figure 3f.

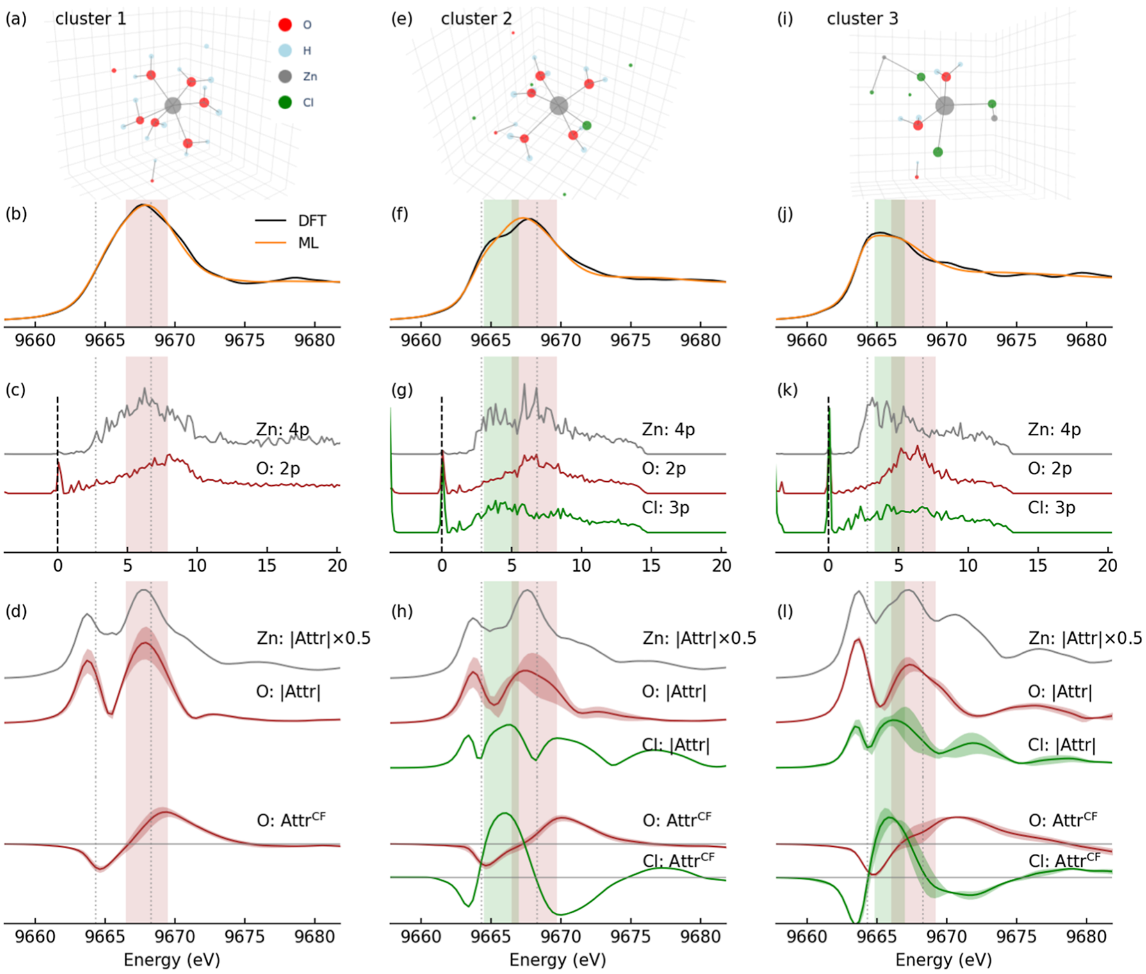


Figure 4. Interpretation of node attributions of the ML model for cluster 1 (a−d), cluster 2 (e−h), and cluster 3 (i−l), with each column corresponding to one cluster. The coordination environments for the clusters are listed in Table 1. (a, e, i) Magnitude of node attributions averaged over all energies, with marker color denoting element species (O: red, H: light blue, Zn: gray, and Cl: green) and marker area proportional to attribution magnitude. (b, f, j) DFT-simulated (black) and ML-predicted (orange) XAS spectra. (c, g, k) p-orbital partial density of states (PDOS) referenced to the Fermi level (black dashed line). The PDOS energy is aligned to the XAS excitation energy scale using the ΔSCF alignment procedure described in the Methods section and a constant energy shift. The Fermi level is set to 0 eV for presentation. (d, h, l) Energy-resolved node attributions, with shaded areas above and below the curves representing one standard deviation of equivalent atoms in the FSS. The top shows attribution magnitudes for Zn, O, and Cl nodes. The bottom shows counterfactual (CF) attributions for O and Cl, where gray horizontal lines mark the zero level.

structural parameters listed in Table 1. In the following sections, we use these six representative samples to interpret the ML model predictions.

3.3. Node Attribution Analysis and Correlation with PDOS

We employ the IG to analyze how structural motifs contribute to the ML-predicted spectrum. At each energy, IG quantifies the sensitivity of the predicted intensity to changes in the input atomic representations. The resulting attributions are interpreted from a physical perspective. Specifically, we examine the correlation of node attribution with the electronic structure based on the partial density of states (PDOS).

Consider the dominant dipole transitions in Zn K-edge X-ray absorption near-edge structure (XANES)

$$\mu(E) \propto |\langle \varphi_c|\hat{\mathbf{e}}\cdot\mathbf{r}|\varphi_{1s}\rangle|^2 \rho_c(E)$$

where $|\varphi_{1s}\rangle$ and $|\varphi_c\rangle$ are Kohn−Sham 1s and conduction band orbitals, $\hat{\mathbf{e}}$ is the polarization direction of the electric field, $\mathbf{r}$ is the electron position operator, and $\rho_c(E)$ is the density of states at excitation energy $E = \varepsilon_c - \varepsilon_{1s}$. Figures 4b,f,j and 5b,f,j show the simulated and ML-predicted XAS spectra for representative structures from the six clusters. Figures 4c,g,k and 5c,g,k show the PDOS of the absorber Zn 4p and the O 2p and Cl 3p states in the FSS. The shape of the unoccupied Zn 4p PDOS closely matches the near-edge structure of the XAS. The Zn 4p states exhibit strong hybridization with the O 2p and Cl 3p states in the FSS, influencing both the peak positions and intensities.

Next, we analyze the node-wise attribution obtained from the trained ML model. The detailed formulation can be found in the Methods section, and here, we briefly summarize it. IG computes the path integral of the below gradient

$$\frac{\partial \mu(E)}{\partial h_i^{(0)}} = \frac{\partial \mu(E)}{\partial h_0^{(L)}} \frac{\partial h_0^{(L)}}{\partial h_i^{(0)}}$$

where $\mu(E)$ is the predicted energy-dependent spectrum intensity, $h_i^{(0)}$ is the 64-dimensional initial embedding depending solely on atomic species, and $h_0^{(L)}$ is the final embedding of the absorbing atom (node 0) after L message-

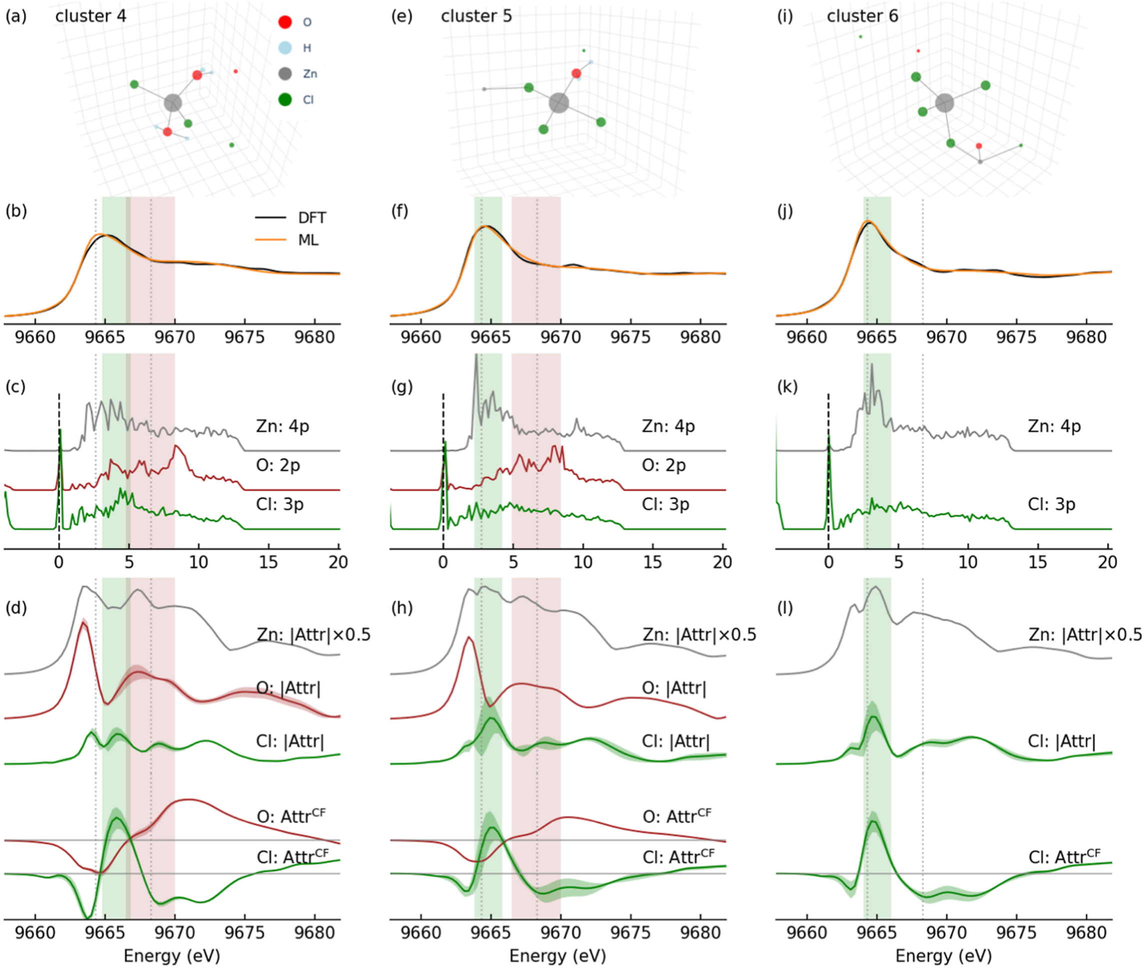


Figure 5. Interpretation of node attributions of the ML model for cluster 4 (a−d), cluster 5 (e−h), and cluster 6 (i−l), with each column corresponding to one cluster. The coordination environments for the clusters are listed in Table 1. (a, e, i) Magnitude of node attributions averaged over all energies, with marker color denoting element species (O: red, H: light blue, Zn: gray, and Cl: green), and marker area proportional to attribution magnitude. (b, f, j) DFT simulated (black) and ML predicted (orange) XAS spectra. (c, g, k) p-orbital partial density of states (PDOS) referenced to the Fermi level (black dashed line). The PDOS energy is aligned to the XAS excitation energy scale using the ΔSCF alignment procedure described in the Methods section and a constant energy shift. The Fermi level is set to 0 eV for presentation. (d, h, l) Energy-resolved node attributions, with shaded areas above and below the curves representing one standard deviation of equivalent atoms in the FSS. The top shows attribution magnitudes for Zn, O, and Cl nodes. The bottom shows counterfactual (CF) attributions for O and Cl, where gray horizontal lines mark the zero level.

passing layers. On the right-hand side, the first derivative term represents the gradient propagation through the spectrum head, and the second Jacobian term represents the propagation through the message-passing layers of the GNN. Together, this gradient encodes the structural dependence of the spectrum on the atomic configuration.

For each of the 64 dimensions of $h_i^{(0)}$, a gradient integral $\mathrm{IG}_{i,d}(E)$ ($d = 1-64$) is obtained by integrating the gradient along a linear path from baseline embedding to the actual node embedding. Here the baseline h^{base} is chosen to be a zero vector representing the absence of structural information. Because the signs in node embedding vectors do not have physical meaning unless a counterfactual (CF) analysis is performed (introduced later in this section), the importance of node i at energy E is quantified by the L_1-norm of its integrated gradients

$$|\mathrm{Attr}_i(E)| = \|\mathrm{IG}_{i,\mathrm{d}}(E)\|_1 = \sum_{d=1}^{64} |\mathrm{IG}_{i,\mathrm{d}}(E)|$$

$|\mathrm{Attr}_i(E)|$ represents the energy-dependent sensitivity of the predicted intensity to perturbations in the atomic representations.

In Figures 4a,e,i and 5a,e,i, the marker size represents the average node importance for representative structures from clusters 1−6, obtained by averaging the attribution over all 200 energy points. Atoms in the FSS contribute most significantly to the spectrum prediction, with moderate contributions from a few second shell atoms within 5 Å of the absorber. For clarity, only atoms of which the importance exceeds 10% of the mean importance of O and Cl in the FSS are displayed. Plots with thresholds set to 5% and 20% are shown in Figures S7 and S8, respectively. This spatial localization of attribution is consistent with the local nature of XANES and validates our choice of using only the absorber features for spectral prediction. In structures containing Zn−Cl−Zn bridging motifs, second-shell Zn atoms contribute non-negligibly to the prediction, indicating that the model captures the second-shell Zn coupling mediated through the bridging Cl ligands.

Table 2. Summary of Gradient-Based Attribution Analyses, Results, and Physical Interpretations

attribution scheme	attribution	observation	interpretation
Node attribution: Gradient: $\frac{\partial \boldsymbol{\mu}}{\partial h^{(\mathbf{0})}}$; Baseline: zero vector	$\lvert \mathrm{Attr}_{\mathrm{Zn}} \rvert$	Distinct attribution patterns for 4-, 5-, and 6-coordinated FSS	GNN differentiates the local geometry around the absorber
	$\lvert \mathrm{Attr}_{\mathrm{O}} \rvert$	• 6-coord: O attributions strongly correlated with Zn attribution shape and white-line peak	Attribution aligns with Zn 4p−O 2p hybridization features in PDOS; GNN captures the structural origin of spectrum intensity
		• 4-coord: Secondary contribution; modify XAS shape	
	$\lvert \mathrm{Attr}_{\mathrm{Cl}} \rvert$	• 4-coord: Cl attributions strongly correlated with Zn attribution shape and white-line peak	Attribution aligns with Zn 4p−Cl 3p hybridization features in PDOS; GNN captures the structural origin of spectrum intensity
		• 6-coord: Secondary contribution; modify XAS shape	
CF attribution: Gradient: $\frac{\partial \boldsymbol{\mu}}{\partial h^{(\mathbf{0})}}$; Baseline: $(h_{\mathrm{O}}{}^{(0)} + h_{\mathrm{Cl}}^{(0)})/2$	$\mathrm{Attr}_{\mathrm{O}}^{\mathrm{CF}}$	Positive near 9670 eV: signature of O ligand presence in tetrahedral environments	GNN encodes ligand-specific features and relative position of O 2p and Cl 3p conduction bands
	$\mathrm{Attr}_{\mathrm{Cl}}^{\mathrm{CF}}$	Positive near 9666 eV: signature of Cl ligand presence in octahedral environments	
Bond-length attribution: Gradient: $\frac{\partial \boldsymbol{\mu}}{\partial R}$; Baseline: R + 0.05 Å	$\mathrm{Attr}^{\mathrm{bond\_len}}$	Elongating Zn−ligand bonds shifts spectral features toward lower energy	Interpretation with the multiple-scattering theory: increasing bond length lowers constructive interference energy
		White-line intensity increases when Zn-ligand bonds become more uniform	Symmetric first-shell geometry enhances constructive scattering
		Zn−O and Zn−Cl attributions show out-of-phase oscillations	Oscillating pattern is influenced by the phase differences between Zn−O and Zn−Cl scattering paths

Figures 4d,h,i and 5d,h,i present the energy-resolved attributions for the absorbing Zn and its first-shell O and Cl ligands. The Zn attributions exhibit four characteristic peaks between 9662 and 9674 eV, whose relative intensities vary with coordination geometry. Clusters 1 and 2 (octahedral Zn-$(H_2O)_6$ and $Zn(H_2O)_5Cl$) show dominant peaks at 9664 and 9668 eV, while clusters 4−6 (tetrahedral $ZnCl_4$-type) show four peaks of comparable height, with the second peak (∼9665 eV) increasing in intensity as the number of Cl neighbors increases. Cluster 3 (pentahedral) shows intermediate features between the octahedral and tetrahedral limits. These trends demonstrate that GNN differentiates the local geometry of the absorber. Through the message-passing layers, the embedding of central Zn is modulated by the ligand-field symmetry and propagated to the final spectral prediction.

In the octahedral clusters 1 and 2, the O-ligand attributions resemble the Zn attributions, both exhibiting peaks at about 9664 and 9668 eV. This correlation indicates that information from the ligands forms a major component of the message passing that shapes the Zn embedding. The PDOS in Figures 4c,g,k and 5c,g,k confirms that Zn 4p and O 2p states hybridize in this energy region, giving rise to the white-line peaks. Similarly, for the tetrahedral clusters 4, 5 and 6, the Cl attributions coincide with the second Zn-attribution peak, showing a strong peak at about 9665 eV. The same energy region corresponds to strong Zn 4p-Cl 3p hybridization, highlighted by the green-shaded region in the PDOS plots. For the nondominant species (Cl in octahedral and O in tetrahedral environments), node attribution peak positions are also correlated with hybridization energies, e.g., the 9666 eV Zn−Cl feature in cluster 2 and the 9668 eV Zn−O feature in clusters 4 and 5. Cluster 3, which contains two O and three Cl ligands (distorted pentahedral geometry), exhibits a mixed behavior, with both Cl and O attributions contribute significantly.

These observations demonstrate that the ML model captures the ligand contributions to the spectral intensity, consistent with the Zn-ligand hybridizations revealed by the PDOS, which determine the XAS spectral shape. Node attribution and PDOS analysis explain the XAS features in different aspects: Node attribution captures the structural origin of the intensity, while PDOS reveals its electronic origin. Thus, the ML model reproduces the physical correlation between the ligand environment and spectral features.

The attribution magnitude quantifies how strongly each atom influences the prediction but does not indicate whether that influence is positive or negative. To address questions such as how swapping a Cl ligand for an O ligand would affect the spectrum shape, a more straightforward explanation is necessary. We compute a CF attribution for ligand atoms O and Cl. Unlike the standard node-wise attribution that uses a zero baseline, CF uses the average of the O and Cl node embeddings, $h^{\mathrm{base}} = \frac{1}{2}(h_{\mathrm{O}}^{(0)} + h_{\mathrm{Cl}}^{(0)})$, as a chemically neutral reference. Because this baseline is between two real chemical environments, the sign of the CF attribution becomes physically interpretable. The CF attribution is calculated as $\mathrm{Attr}_i^{\mathrm{CF}}(E) = \sum_{d=1}^{64} \mathrm{IG}_{i,d}(E)$. For the original Cl ligand, a positive CF value indicates that replacing O with Cl increases the spectral intensity at that energy, while a negative value indicates the opposite. It should be noted that although CF attribution provides directional information, it can only be defined for two species. For broader comparisons across all atoms, magnitude-based attribution with a zero baseline remains the most general measure. A summary comparing the node-wise attribution, CF attribution, as well as the bond-length attributions in the next section is in Table 2.

The CF attributions for O and Cl have similar patterns across all of the clusters. The CF attribution of O shows a positive peak between 9669 and 9671 eV and a negative region near 9665 eV. The trend of Cl is reversed, with positive contributions around 9665 to 9666 eV and negative values between 9668 and 9672 eV. This trend can be understood based on the electronic structure analysis. In the $ZnCl_2$ solution, the valence band maximum is dominated by the Cl 3p states, and the water $1b_1$ lone-pair states are pushed to lower energy and become chemically more stable.[29,60] As a result, the Cl 3p antibonding states are lower in energy than the O 2p antibonding states in the conduction band.

These opposite trends reflect the substitution relationship between the two ligand species and can be used to identify the spectroscopic signatures of the O-coordinated or Cl-coordinated environments. In octahedral cluster 2, the Cl CF attribution displays a positive peak at 9666 eV, marking the

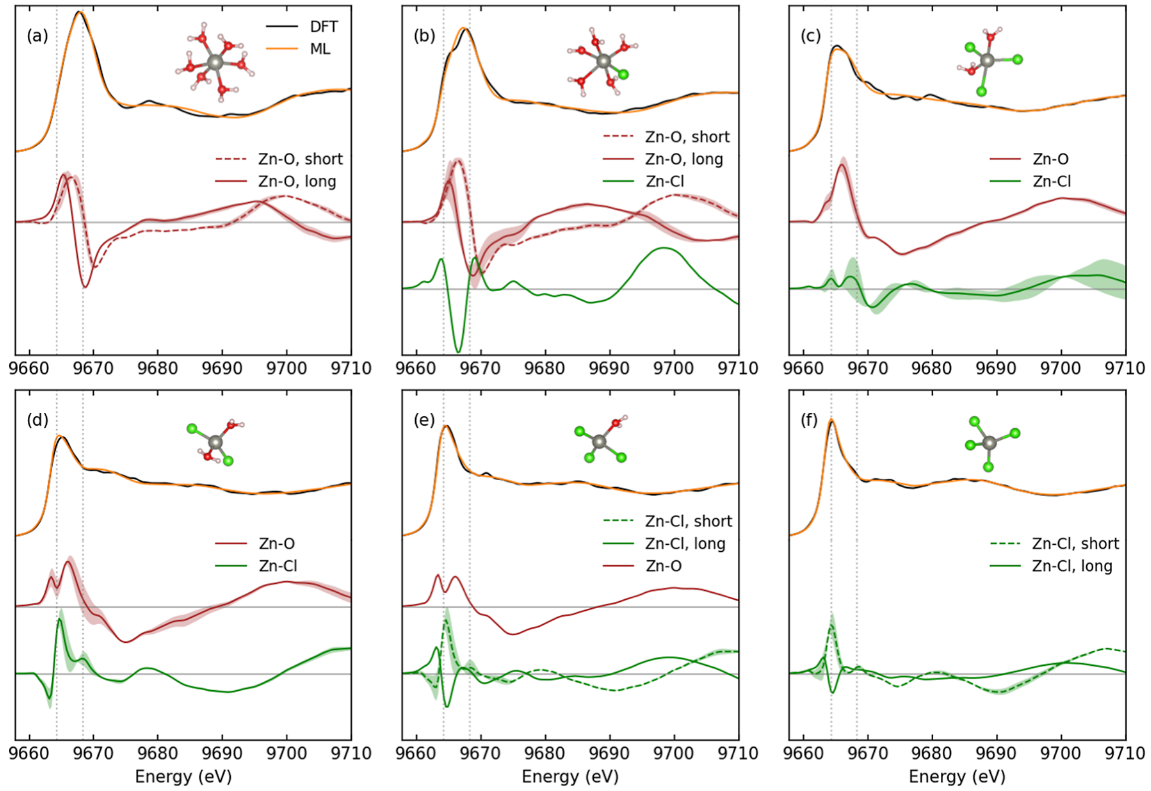


Figure 6. Bond-length attributions for six representative local motifs: (a) $Zn(H_2O)_6$, (b) $Zn(H_2O)_5Cl$, (c) $Zn(H_2O)_2Cl_3$, (d) $Zn(H_2O)_2Cl_2$, (e) $Zn(H_2O)Cl_3$, and (f) $ZnCl_4$. These are the representative structures from the six clusters listed in Table 1. The inset atomic structures show the absorbing Zn site and the Cl atoms and water molecule within the first shell. The top of each subplot shows the comparison of simulated (black) and ML-predicted (orange) XAS. Red and green curves show the bond-length attributions for Zn−O and Zn−Cl bonds, respectively; solid and dashed lines correspond to longer and shorter bonds within each motif. Gray horizontal lines indicate the zero-attribution level. In (a,b), the average lengths for long and short Zn−O bonds are 2.28 Å and 2.05 Å, respectively. In (e) and (f), the average lengths for long and short Zn−Cl bonds are 2.42 Å and 2.24 Å, respectively. Positive attribution corresponds to an increase in spectral intensity upon elongation of the bond, whereas negative attribution corresponds to a decrease.

signature of partial chlorination in an otherwise fully hydrated FSS. The simulated spectrum exhibits a shoulder at this energy (Figure 4f). Although the ML-predicted spectrum does not reproduce a distinct shoulder, the intensity around 9666 eV is elevated compared to cluster 1. In tetrahedral clusters 3 and 4, the O CF attribution has a broad peak at about 9671 eV, indicating that the presence of O rather than Cl increases the intensity in this region. This is consistent with the spectra in Figure 3e, where the bump at about 9671 eV grows with the number of O ligands, indicated by the black arrow. Hence, this feature can serve as a signature of the presence of O ligands in tetrahedral environments. Overall, the CF analysis demonstrates that the GNN model correctly encodes how ligand substitution (i.e., O or Cl) modulates the spectral intensity.

In summary, the gradient-based node attributions quantify the ML model's sensitivity to perturbations in atomic representation. The good alignment between energy-resolved attributions and PDOS analysis demonstrates that the GNN learns the physical structural-spectrum relationships as revealed by Zn-ligand hybridization in the electronic structure.[61,62] The CF analysis further distinguishes whether the presence of a ligand enhances or suppresses the spectral intensity and helps identify characteristic features, such as the Cl signature in octahedral and the O signature in tetrahedral environments. The main findings are summarized in Table 2.

3.4. Bond-Length Attribution Analysis

To investigate how the ML model encodes bond distances, we perform a bond-length attribution analysis using the IG method. For each bond i, IG quantifies the integral of gradient $\frac{\partial \mu(E)}{\partial R_i}$, which measures how the predicted spectrum intensity $\mu(E)$ changes with variations in bond length R_i. The integration is carried out along a continuous path from a baseline of $R_i^{\text{base}} = R_i + \Delta R$, where $\Delta R = 0.05$ Å (i.e., the target bond is elongated by 0.05 Å), to the actual bond length R_i. We take the negative of calculated integrated gradients as the bond-length attribution, which corresponds to the change in spectral intensity associated with an incremental elongation of the specific bond. To assess the robustness of the bond-stretching baseline choice, we performed tests using $\Delta R = 0.01$, 0.03, and 0.07 Å. The results are shown in Figures S9−S11. The qualitative features of bond attributions remain consistent across the tested ΔR values.

We interpret the bond distance attribution trends based on the multiple scattering theory, where the spectral features $\chi(E)$ can be rationalized as a sum of the contributions from different scattering paths.[63] $\chi(E)$ is defined as the spectral contrast of $\mu(E)$ with respect to a smooth atomic background $\mu_0(E)$, $\chi(E) = \frac{\mu(E) - \mu_0(E)}{\Delta\mu_0}$, where $\Delta\mu_0$ is the measured jump in $\mu(E)$ at the absorption edge E_0.[1,64] Specifically, leading contributions from the two-leg paths $\chi_i(E)$ of each ligand i can be simplified according to Sayers and Stern's formulation:[1,65]

$$\chi_i(E) \propto f_i(k)\sin(2kR_i + 2\eta_i(k)) \qquad (1)$$

where $f_i(k)$ is the effective scattering amplitude of the ligand atom i, R_i is the absorber−ligand distance, $\eta_i(k)$ is the phase shift, and $k = \sqrt{2m_e(E - E_0)}/\hbar$ is the photoelectron wavenumber corresponding to excitation energy E relative to E_0. The $2kR_i$ term determines the oscillation in k space. Variations in bond length R_i change the oscillation frequency of $\chi_i(E)$ and thus the interference pattern among scattering paths, leading to energy-dependent bond-length attributions.

Figure 6 shows the bond-length attributions for representative spectra from the six clusters. For the fully hydrated $Zn(H_2O)_6$ shown in Figure 6a, the six Zn−O bonds are grouped into two sets: a shorter group with an average bond length of 2.05 Å and a longer group with an average length of 2.22 Å. According to eq 1, a longer bond length increases the oscillation frequency of $\chi_i(E)$ and shifts the constructive-interference condition to slightly lower energies. This is reflected in the bond-length attribution, where stretching the Zn−O bonds shifts the scattering peak to the lower-energy side of the white-line peak, and the peak position for the longer bonds appears at a lower energy.

When the FSS is asymmetric, e.g., distorted $Zn(H_2O)_6$ with different Zn−O bond lengths, stretching the shorter Zn−O bonds slightly increases the white-line intensity, whereas stretching the longer bonds reduces it. This indicates that the white-line peak is enhanced when the first-shell environment becomes more symmetric. The tetrahedral $ZnCl_4$ complex (Figure 6f) follows the same trend in which more symmetric motifs exhibit larger white-line intensity. In mixed-ligand motifs, such as $Zn(H_2O)_5Cl$ and $Zn(H_2O)Cl_3$, the Zn−O and Zn−Cl bond-length attributions follow similar symmetry-dependent responses. According to eq 1, uniform Zn-ligand distances yield equal oscillation phases in the $2kR_i$ term, leading to constructive interference among two-leg scattering paths and a stronger total amplitude. Conversely, structural asymmetry acts as a form of bond disorder, which causes amplitude reduction and is often described by Debye−Waller damping factors in the extended X-ray absorption fine structure. These results demonstrate that the ML model effectively captures local symmetry through its edge and angular features and learns that the white line peak intensity is maximized when dominating scattering paths contribute constructively, e.g., the first coordination shell is geometrically uniform.

For partially hydrated and partially chlorinated complexes, the effects of stretching Zn−O and Zn−Cl bonds are different and often exhibit out-of-phase oscillations. The O and Cl ligands have different scattering amplitudes $f_i(k)$ and oscillation frequencies, resulting in more complex energy-dependent interference between Zn−O and Zn−Cl scattering paths. The XAS fine structure is related to the coherent sum of multiple scattering paths with different phases, $f_O(k)\sin(2kR_O + 2\eta_O(k))$ and $f_{Cl}(k)\sin(2kR_{Cl} + 2\eta_{Cl}(k))$, where the relative phases determine whether the individual path contributions interfere constructively or destructively. Taking $Zn(H_2O)_2Cl_2$ as an example, when Zn−Cl bonds are elongated, their oscillation frequencies increase, shifting the constructive interference to slightly lower photon energies (Figure 6d). Nevertheless, the bond-length attribution behavior for mixed ligands is more complicated than that in single-ligand systems. Despite the difficulty in rationalization, the ML model captures the alternating constructive and destructive interference of Zn−O and Zn−Cl paths and the phase-beating pattern that governs the near-edge structure of mixed-ligand Zn complexes.

4. CONCLUSIONS

We have developed a physics-guided ML model for Zn K-edge XAS for aqueous $ZnCl_2$ solutions across a broad concentration range. ML-predicted spectra agree well with the experimental measurements. Applying post hoc many-body shakeup corrections improves the agreement with experimental data. The ML framework is computationally efficient and scalable, enabling the accurate prediction of XAS for large dilute systems, which are otherwise computationally prohibitive with conventional ab initio methods.

UMAP analysis shows that the learned latent features from GNN accurately capture the local coordination environments of Zn absorbers. Gradient-based attribution analyses reveal that the model captures physically consistent structure-spectrum relationships. Node-wise attributions show that the energy-resolved importance of ligands O and Cl strongly correlates with the orbital hybridization patterns observed in PDOS. CF attribution analysis further clarifies the spectral sensitivity on atom type, identifying signatures of O coordination in tetrahedral $ZnCl_4$-like environments and Cl coordination in octahedral $Zn(H_2O)_6$-like environments. Furthermore, edge-wise attribution analysis correlates bond length perturbations with changes in the scattering paths, consistent with multiple-scattering theory.

Together, these results demonstrate that the ML model can recover physically meaningful principles. By correlating GNN attributions to orbital hybridization and photoelectron scattering, our approach bridges data-driven predictions with a quantum mechanical understanding. This study establishes a physics-guided architecture and a general pathway for interpretable ML in spectroscopy.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.6c00464.

> Spectral functions calculated from configurations containing representative structural motifs; decile plots of machine learning model predictions on test data set; concentrations of $ZnCl_2$ solutions and the numbers of molecules in the MD simulations; root mean square error (RMSE) and integrated absolute error (IAE) between experimentally measured and ML-predicted spectra; O and Cl distributions in the FSS for 0.1 m $ZnCl_2$ solution; evolution trend of XAS spectra with concentrations; UMAP plot of latent features from GNN color coded by second solvation shell atoms; UMAP projections with different initialization seeds; average node importance for representative structures in clusters 1 − 6; and bond-length attributions (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Chuntian Cao − *Artificial Intelligence Department, Brookhaven National Laboratory, Upton, New York 11973,* United States; orcid.org/0000-0003-4753-7334; Email: ccao@bnl.gov

Amy C. Marschilok − *Institute of Sustainability, Electrification, and Energy (I:SEE), Stony Brook University, Stony Brook, New York 11794, United States; Department of Chemistry and Department of Materials Science and Chemical Engineering, Stony Brook University, Stony Brook, New York 11794, United States; Interdisciplinary Science Department, Brookhaven National Laboratory, Upton, New York 11973, United States;* orcid.org/0000-0001-9174-0474; Email: amy.marschilok@stonybrook.edu

Deyu Lu − *Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, United States;* orcid.org/0000-0003-4351-6085; Email: dlu@bnl.gov

Authors

Boyang Li − *Artificial Intelligence Department, Brookhaven National Laboratory, Upton, New York 11973, United States*

Armando Rodriguez Campos − *Institute of Sustainability, Electrification, and Energy (I:SEE), Stony Brook University, Stony Brook, New York 11794, United States; Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; Interdisciplinary Science Department, Brookhaven National Laboratory, Upton, New York 11973, United States;* orcid.org/0000-0003-3050-5926

Alexis Pace − *Institute of Sustainability, Electrification, and Energy (I:SEE), Stony Brook University, Stony Brook, New York 11794, United States;* orcid.org/0009-0008-0095-6834

Joshua J. Kas − *Department of Physics, University of Washington, Seattle, Washington 98195, United States*

Xifan Wu − *Department of Physics, Temple University, Philadelphia, Pennsylvania 19122, United States;* orcid.org/0000-0001-6561-8942

Lu Ma − *National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, New York 11973, United States*

Dali Yang − *National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, New York 11973, United States;* orcid.org/0000-0002-8589-1003

Wei Xu − *Artificial Intelligence Department, Brookhaven National Laboratory, Upton, New York 11973, United States*

Shinjae Yoo − *Artificial Intelligence Department, Brookhaven National Laboratory, Upton, New York 11973, United States*

Esther S. Takeuchi − *Institute of Sustainability, Electrification, and Energy (I:SEE), Stony Brook University, Stony Brook, New York 11794, United States; Department of Chemistry and Department of Materials Science and Chemical Engineering, Stony Brook University, Stony Brook, New York 11794, United States; Interdisciplinary Science Department, Brookhaven National Laboratory, Upton, New York 11973, United States;* orcid.org/0000-0001-8518-1047

Kenneth J. Takeuchi − *Institute of Sustainability, Electrification, and Energy (I:SEE), Stony Brook University, Stony Brook, New York 11794, United States; Department of Chemistry and Department of Materials Science and Chemical Engineering, Stony Brook University, Stony Brook, New York 11794, United States; Interdisciplinary Science Department, Brookhaven National Laboratory, Upton, New York 11973, United States;* orcid.org/0000-0001-8129-444X

Shan Yan − *Institute of Sustainability, Electrification, and Energy (I:SEE), Stony Brook University, Stony Brook, New York 11794, United States; Interdisciplinary Science Department, Brookhaven National Laboratory, Upton, New York 11973, United States;* orcid.org/0000-0002-9715-9100

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.6c00464

Notes

Data and Code Availability: The data used to train the machine learning model and the code used in this work are openly available at doi.org/10.5281/zenodo.19684914.

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, through Contract No. DE-SC0012704, under the Chemical and Materials Sciences to Advance Clean-Energy Technologies and Transform Manufacturing (CEM) program. This research used theory and computational resources at the Center for Functional Nanomaterials and beamline 7-BM, Quick X-ray Absorption and Scattering (QAS) of the National Synchrotron Light Source II, which are U.S. Department of Energy (DOE) Office of Science User Facilities operated for the DOE Office of Science by Brookhaven National Laboratory under contract No. DE-SC0012704. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy User Facility using NERSC awards ERCAP0030552 and ERCAP0032397. X.W. acknowledges support from the National Science Foundation through Grant No. DMR-2053195 for their intellectual input to deep potential molecular dynamics of $ZnCl_2$ solutions. J.J.K. acknowledges support from the Theory Institute for Materials and Energy Spectroscopies (TIMES) at SLAC, funded by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering under Contract No. DE-AC02-76SF00515, for their contribution to spectral function calculation and many-body shakeup correction. We thank Dr. Shubha R. Kharel for the fruitful discussions. E.S.T. acknowledges support as the William and Jane Knapp Chair in Energy and the Environment.

■ REFERENCES

(1) Rehr, J. J.; Albers, R. C. Theoretical approaches to x-ray absorption fine structure. *Rev. Mod. Phys.* **2000**, *72* (3), 621−654.

(2) Penner-Hahn, J. E. X-ray absorption spectroscopy in coordination chemistry. *Coord. Chem. Rev.* **1999**, *190−192*, 1101−1123.

(3) Yano, J.; Yachandra, V. K. X-ray absorption spectroscopy. *Photosynth. Res.* **2009**, *102* (2), 241−254.

(4) Wang, Z.; Kuang, J.; Rodriguez-Campos, A.; Cao, C.; Kingan, A.; Barry, P. J.; Hill, R. C.; Arnot, D. J.; Christianne, A.; Bock, D. C.; Du, Y.; Bak, S. M.; Ma, L.; Yang, D.; Tayal, A.; Drakopoulos, M.; Zhong, Z.; Vo, N. T.; Kisslinger, K.; Tong, X.; Takeuchi, E. S.; Carbone, M. R.; Lu, D.; Wang, L.; Yan, S.; Takeuchi, K. J.; Marschilok, A. C. Elucidating the Discharge Behavior of Aqueous Zinc Sulfur Batteries in the Presence of Molybdenum(IV) Chalcogenide Catalyst: The Criticality of Interfacial Electrochemistry. *ACS Appl. Mater. Interfaces* **2024**, *16* (49), 67730−67742.

(5) Cao, C.; Carbone, M. R.; Komurcuoglu, C.; Shekhawat, J. S.; Sun, K.; Guo, H.; Liu, S.; Chen, K.; Bak, S.-M.; Du, Y.; Weiland, C.; Tong, X.; Steingart, D. A.; Yoo, S.; Artrith, N.; Urban, A.; Lu, D.; Wang, F. Atomic insights into the oxidative degradation mechanisms of sulfide solid electrolytes. *Cell Rep. Phys. Sci.* **2024**, *5* (4), 101909.

(6) Kwon, H.; Hsu, T.; Sun, W.; Jeong, W.; Aydin, F.; Chapman, J.; Chen, X.; Lordi, V.; Carbone, M. R.; Lu, D.; Zhou, F.; Anh Pham, T. Spectroscopy-guided discovery of three-dimensional structures of disordered materials with diffusion models. *Machine Learning: Science and Technology* **2024**, *5* (4), 045037.

(7) Na Narong, T.; Zachko, Z. N.; Torrisi, S. B.; Billinge, S. J. L. Interpretable multimodal machine learning analysis of X-ray absorption near-edge spectra and pair distribution functions. *npj Comput. Mater.* **2025**, *11* (1), 98.

(8) Kharel, S. R.; Meng, F.; Qu, X.; Carbone, M. R.; Lu, D. OmniXAS: A universal deep-learning framework for materials x-ray absorption spectra. *Phys. Rev. Mater.* **2025**, *9* (4), 043803.

(9) Kotobi, A.; Singh, K.; Hoche, D.; Bari, S.; Meissner, R. H.; Bande, A. Integrating Explainability into Graph Neural Network Models for the Prediction of X-ray Absorption Spectra. *J. Am. Chem. Soc.* **2023**, *145* (41), 22584−22598.

(10) Liang, Z.; Carbone, M. R.; Chen, W.; Meng, F.; Stavitski, E.; Lu, D.; Hybertsen, M. S.; Qu, X. Decoding structure-spectrum relationships with physically organized latent spaces. *Phys. Rev. Mater.* **2023**, *7* (5), 053802.

(11) Tetef, S.; Govind, N.; Seidler, G. T. Unsupervised machine learning for unbiased chemical classification in X-ray absorption spectroscopy and X-ray emission spectroscopy. *Phys. Chem. Chem. Phys.* **2021**, *23* (41), 23586−23601.

(12) Chen, Y.; Chen, C.; Hwang, I.; Davis, M. J.; Yang, W.; Sun, C.; Lee, G.-H.; McReynolds, D.; Allan, D.; Marulanda Arias, J.; Ong, S. P.; Chan, M. K. Y. Robust Machine Learning Inference from X-ray Absorption Near Edge Spectra through Featurization. *Chem. Mater.* **2024**, *36* (5), 2304−2313.

(13) Rankine, C. D.; Penfold, T. J. Accurate, affordable, and generalizable machine learning simulations of transition metal x-ray absorption spectra using the XANESNET deep neural network. *J. Chem. Phys.* **2022**, *156* (16), 164102.

(14) Jeong, W.; Sun, W.; Calegari Andrade, M. F.; Wan, L. F.; Willey, T. M.; Nielsen, M. H.; Pham, T. A. Integrating Machine Learning Potential and X-ray Absorption Spectroscopy for Predicting the Chemical Speciation of Disordered Carbon Nitrides. *Chem. Mater.* **2024**, *36* (9), 4144−4156.

(15) Chen, Z.; Andrejevic, N.; Drucker, N. C.; Nguyen, T.; Xian, R. P.; Smidt, T.; Wang, Y.; Ernstorfer, R.; Tennant, D. A.; Chan, M.; Li, M. Machine learning on neutron and x-ray scattering and spectroscopies. *Chem. Phys. Rev.* **2021**, *2* (3), 031301.

(16) Carbone, M. R.; Topsakal, M.; Lu, D.; Yoo, S. Machine-Learning X-Ray Absorption Spectra to Quantitative Accuracy. *Phys. Rev. Lett.* **2020**, *124* (15), 156401.

(17) Ghose, A.; Segal, M.; Meng, F.; Liang, Z.; Hybertsen, M. S.; Qu, X.; Stavitski, E.; Yoo, S.; Lu, D.; Carbone, M. R. Uncertainty-aware predictions of molecular x-ray absorption spectra using neural network ensembles. *Phys. Rev. Res.* **2023**, *5* (1), 013180.

(18) Gleason, S. P.; Carbone, M. R.; Lu, D.; Ciston, J. CuXASNet: Rapid and accurate prediction of copper L-edge x-ray absorption spectra using machine learning. *Phys. Rev. Mater.* **2025**, *9* (7), 073803.

(19) Zhou, H.; Xu, L.; Ren, Z.; Zhu, J.; Lee, C. Machine learning-augmented surface-enhanced spectroscopy toward next-generation molecular diagnostics. *Nanoscale Adv.* **2023**, *5* (3), 538−570.

(20) Sabattini, L.; Coriolano, A.; Casert, C.; Forti, S.; Barnard, E. S.; Beltram, F.; Pontil, M.; Whitelam, S.; Coletti, C.; Rossi, A. Towards AI-driven autonomous growth of 2D materials based on a graphene case study. *Commun. Phys.* **2025**, *8* (1), 180.

(21) Westermayr, J.; Marquetand, P. Machine learning spectroscopy to advance computation and analysis. *Chem. Sci.* **2025**, *16* (46), 21660−21676.

(22) Xu, Y.; Bian, D.; Ju, C.-W.; Zhao, F.; Xie, P.; Wang, Y.; Hu, W.; Sun, Z.; Zhang, J. Z. H.; Zhu, T. Pretrained E(3)-equivariant message-passing neural networks with multi-level representations for organic molecule spectra prediction. *npj Comput. Mater.* **2025**, *11* (1), 203.

(23) Kim, T.; Lee, D. Unveiling key descriptors for electrical resistivity of alloys using high-throughput experiments and explainable AI. *npj Comput. Mater.* **2025**, *11* (1), 300.

(24) Oviedo, F.; Ferres, J. L.; Buonassisi, T.; Butler, K. T. Interpretable and Explainable Machine Learning for Materials Science and Chemistry. *Acc. Mater. Res.* **2022**, *3* (6), 597−607.

(25) Zhou, R.; Bao, L.; Bu, W.; Zhou, F. Exploring high-performance viscosity index improver polymers via high-throughput molecular dynamics and explainable AI. *npj Comput. Mater.* **2025**, *11* (1), 52.

(26) Ji, X. A perspective of ZnCl2 electrolytes: The physical and electrochemical properties. *eScience* **2021**, *1* (2), 99−107.

(27) Zhu, Y.; Yin, J.; Zheng, X.; Emwas, A.-H.; Lei, Y.; Mohammed, O. F.; Cui, Y.; Alshareef, H. N. Concentrated dual-cation electrolyte strategy for aqueous zinc-ion batteries. *Energy Environ. Sci.* **2021**, *14* (8), 4463−4473.

(28) Dhakal, D.; Driscoll, D. M.; Govind, N.; Stack, A. G.; Rampal, N.; Schenter, G.; Mundy, C. J.; Fister, T. T.; Fulton, J. L.; Balasubramanian, M.; Seidler, G. T. The evolution of solvation symmetry and composition in Zn halide aqueous solutions from dilute to extreme concentrations. *Phys. Chem. Chem. Phys.* **2023**, *25* (34), 22650−22661.

(29) Cao, C.; Kingan, A.; Hill, R. C.; Kuang, J.; Wang, L.; Zhang, C.; Carbone, M. R.; van Dam, H.; Yoo, S.; Yan, S.; Takeuchi, E. S.; Takeuchi, K. J.; Wu, X.; Abeykoon, A. M. M.; Marschilok, A. C.; Lu, D. Resolving the Solvation Structure and Transport Properties of Aqueous Zinc Electrolytes from Salt-in-Water to Water-in-Salt Using Neural Network Potential. *PRX Energy* **2025**, *4* (2), 023004.

(30) McInnes, L.; Healy, J.; Melville, J. U. Uniform manifold approximation and projection for dimension reduction. *arXiv* **2020**, arXiv:1802.03426.

(31) Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic Attribution for Deep Networks. *Proceedings of the 34th International Conference on Machine Learning, PMLR: Proceedings of Machine Learning Research* 2017, *70*, 3319−3328.

(32) Mudrakarta, P. K.; Taly, A.; Sundararajan, M.; Dhamdhere, K. In Did the model understand the question?, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers)*, 2018; pp 1896−1906.

(33) Ravel, B.; Newville, M.; Athena, A. R. T. E. M. I. S. HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT. *J. Synchrotron Radiat.* **2005**, *12* (4), 537−541.

(34) Driscoll, D. M.; Lavan, S. N.; Zorko, M.; Redfern, P. C.; Ilic, S.; Agarwal, G.; Fister, T. T.; Assary, R. S.; Cheng, L.; Strmcnik, D.; et al. Emergent solvation phenomena in non-aqueous electrolytes with multiple anions. *Chem* **2023**, *9* (7), 1955−1971.

(35) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter* **2009**, *21* (39), 395502.

(36) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Buongiorno Nardelli, M.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M.; Colonna, N.; Carnimeo, I.; Dal Corso, A.; de Gironcoli, S.; Delugas, P.; DiStasio, R. A.; Ferretti, A.; Floris, A.; Fratesi, G.; Fugallo, G.; Gebauer, R.; Gerstmann, U.; Giustino, F.; Gorni, T.; Jia, J.; Kawamura, M.; Ko, H. Y.; Kokalj, A.; Küçükbenli, E.; Lazzeri, M.; Marsili, M.; Marzari, N.; Mauri, F.; Nguyen, N. L.; Nguyen, H. V.; Otero-de-la-Roza, A.; Paulatto, L.; Poncé, S.; Rocca, D.; Sabatini, R.; Santra, B.; Schlipf, M.; Seitsonen, A. P.; Smogunov, A.; Timrov, I.; Thonhauser, T.; Umari, P.; Vast, N.; Wu, X.; Baroni, S. Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys.: Condens. Matter* **2017**, *29* (46), 465901.

(37) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115* (3), 036402.

https://doi.org/10.1021/acs.jpcb.6c00464

(38) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), 143001.

(39) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.

(40) Chen, M.; Ko, H. Y.; Remsing, R. C.; Calegari Andrade, M. F.; Santra, B.; Sun, Z.; Selloni, A.; Car, R.; Klein, M. L.; Perdew, J. P.; Wu, X. Ab initio theory and modeling of water. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (41), 10846−10851.

(41) Morrone, J. A.; Car, R. Nuclear Quantum Effects in Water. *Phys. Rev. Lett.* **2008**, *101* (1), 017801.

(42) Karsai, F.; Humer, M.; Flage-Larsen, E.; Blaha, P.; Kresse, G. Effects of electron-phonon coupling on absorption spectrum: K edge of hexagonal boron nitride. *Phys. Rev. B* **2018**, *98* (23), 235205.

(43) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54* (16), 11169−11186.

(44) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6* (1), 15−50.

(45) Carbone, M. R.; Meng, F.; Vorwerk, C.; Maurer, B.; Peschel, F.; Qu, X.; Stavitski, E.; Draxl, C.; Vinson, J.; Lu, D. Lightshow: a Python package for generating computational x-ray absorption spectroscopy input files. *Journal of Open Source Software* **2023**, *8*, 5182.

(46) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865−3868.

(47) Prendergast, D.; Galli, G. X-Ray Absorption Spectra of Water from First Principles Calculations. *Phys. Rev. Lett.* **2006**, *96* (21), 215502.

(48) Krause, M. O.; Oliver, J. H. Natural widths of atomic K and L levels, Kα X-ray lines and several KLL Auger lines. *J. Phys. Chem. Ref. Data* **1979**, *8* (2), 329−338.

(49) England, A. H.; Duffin, A. M.; Schwartz, C. P.; Uejio, J. S.; Prendergast, D.; Saykally, R. J. On the hydration and hydrolysis of carbon dioxide. *Chem. Phys. Lett.* **2011**, *514* (4), 187−195.

(50) Woicik, J. C.; Weiland, C.; Jaye, C.; Fischer, D. A.; Rumaiz, A. K.; Shirley, E. L.; Kas, J. J.; Rehr, J. J. Charge-transfer satellites and chemical bonding in photoemission and x-ray absorption of SrTiO3 and rutile TiO2: Experiment and first-principles theory with general application to spectroscopic analysis. *Phys. Rev. B* **2020**, *101* (24), 245119.

(51) Kas, J. J.; Rehr, J. J.; Curtis, J. B. Particle-hole cumulant approach for inelastic losses in x-ray spectra. *Phys. Rev. B* **2016**, *94* (3), 035156.

(52) Woicik, J. C.; Weiland, C.; Rumaiz, A. K.; Brumbach, M. T.; Ablett, J. M.; Shirley, E. L.; Kas, J. J.; Rehr, J. J. Core hole processes in x-ray absorption and photoemission by resonant Auger-electron spectroscopy and first-principles theory. *Phys. Rev. B* **2020**, *101* (24), 245105.

(53) Calandra, M.; Rueff, J. P.; Gougoussis, C.; Céolin, D.; Gorgoi, M.; Benedetti, S.; Torelli, P.; Shukla, A.; Chandesris, D.; Brouder, C. K-edge x-ray absorption spectra in transition-metal oxides beyond the single-particle approximation: Shake-up many-body effects. *Phys. Rev. B* **2012**, *86* (16), 165102.

(54) Guzzo, M.; Lani, G.; Sottile, F.; Romaniello, P.; Gatti, M.; Kas, J. J.; Rehr, J. J.; Silly, M. G.; Sirotti, F.; Reining, L. Valence Electron Photoemission Spectrum of Semiconductors: Ab Initio Description of Multiple Satellites. *Phys. Rev. Lett.* **2011**, *107* (16), 166401.

(55) Kas, J. J.; Rehr, J. J.; Reining, L. Cumulant expansion of the retarded one-electron Green function. *Phys. Rev. B* **2014**, *90* (8), 085112.

(56) Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2022**, *2* (11), 718−728.

(57) Mudrakarta, P. K.; Taly, A.; Sundararajan, M.; Dhamdhere, K. Did the model understand the question? *arXiv* **2018**, arXiv:1805.05492.

(58) Kokhlikyan, N.; Miglani, V.; Martin, M.; Wang, E.; Alsallakh, B.; Reynolds, J.; Melnikov, A.; Kliushkina, N.; Araya, C.; Yan, S. C. A unified and generic model interpretability library for pytorch. *arXiv* **2020**, arXiv:2009.07896.

(59) K-Means clustering, Scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html (accessed 12 09 2025).

(60) Tang, F.; Xu, J.; Qiu, D. Y.; Wu, X. Nuclear quantum effects on the quasiparticle properties of the chloride anion aqueous solution within the GW approximation. *Phys. Rev. B* **2021**, *104* (3), 035117.

(61) Tang, F.; Shi, K.; Wu, X. Exploring the impact of ions on oxygen K-edge X-ray absorption spectroscopy in NaCl solution using the GW-Bethe-Salpeter-equation approach. *J. Chem. Phys.* **2023**, *159* (17), 174501.

(62) Tang, F.; Jiang, X.; Ko, H.-Y.; Xu, J.; Topsakal, M.; Hao, G.; N'Diaye, A. T.; Dowben, P. A.; Lu, D.; Xu, X.; Wu, X. Probing ferroelectricity by x-ray absorption spectroscopy in molecular crystals. *Phys. Rev. Mater.* **2020**, *4* (3), 034401.

(63) Rehr, J. J.; Kas, J. J.; Vila, F. D.; Prange, M. P.; Jorissen, K. Parameter-free calculations of X-ray spectra with FEFF9. *Phys. Chem. Chem. Phys.* **2010**, *12* (21), 5503−5513.

(64) Newville, M. Fundamentals of XAFS. *Reviews in Mineralogy and Geochemistry* **2014**, *78*, 33−74.

(65) Sayers, D. E.; Stern, E. A.; Lytle, F. W. New Technique for Investigating Noncrystalline Structures: Fourier Analysis of the Extended X-Ray Absorption Fine Structure. *Phys. Rev. Lett.* **1971**, *27* (18), 1204−1207.